PURPOSE AND GOALS

The Arizona Electronic Records Research Initiative (Initiative) is a collaboration of six government agencies that brings unique expertise to the challenge of preserving electronic records. The Arizona State Library, Archives and Public Records (ASLAPR), the Arizona Government Information Technology Agency (GITA), and the Arizona Secretary of State (SoS) are each charged by state statute with regulating and managing electronic records. The Arizona Administrative Office of the Courts (AOC) and the Clerk of the Arizona Superior Court in Maricopa County (CASCMC) are concerned that their records comply with standards for electronic records being developing in the legal community. Arizona State University (ASU) has been a leader in investigating student records in electronic format.

These agencies recognize that each has important knowledge to bring to the problem of preserving electronic records, but that none has all the knowledge. They have established a collaborative effort to build on their shared expertise to investigate how to store, manage, and authenticate electronic records in an economical manner and how to ensure that current and future citizens will have the same confidence in the reliability and authenticity of electronic records that they currently have in traditional, paper-based records.

The Initiative will develop and test policies, procedures, and techniques to manage electronic records during their entire life cycle to ensure the reliability and authenticity of those records. The Initiative anticipates successive projects to fully implement a complete method to manage electronic record life cycles. Over time, the Initiative's projects will

Evaluate techniques for preserving access to the information independent of proprietary software used to create the records.

How can state agencies cope with the rapid obsolescence of hardware, software, and data formats? Although the raw data may be physically preserved in a readable format, the operating system and application software necessary to view or duplicate the record may not be available. This project will explore development and use of an open-standard, platform-neutral format to preserve access to the meaningful content of electronic records and the presentation format of that content.

Evaluate techniques to ensure the records remain reliable and trustworthy.

Because these policies, procedures, and techniques may create new versions of electronic records, which differ from their original encoding formats to a standard format to support enduring access, it is essential to develop policies and tools that ensure the new versions are reliable, trustworthy, and acceptable as legal evidence.

Identify administrative barriers to preserving and managing electronic records.

Robust policies, procedures, and techniques to preserve electronic records will be meaningless if records creators and custodians ignore them. They must be easy to use and maintain, and agencies will need adequate resources and training to use them.

To accomplish the above, this first, proposed project of the Initiative will undertake the following goals to investigate the use of a non-proprietary, standard file format to simplify management and minimize the impact of technological obsolescence. It will establish policies and procedures to effectively manage electronic records and ensure their authenticity. It will also identify training, resources, and services that state agencies need to implement those policies and procedures.

1. Test data conversion and migration of documents to a standard format that can be read by a platform-independent viewer.
   * Establish criteria for evaluating the authenticity of an electronic record in terms of content and presentation.
   * Create a standard XML schema for describing and organizing several record sets in diverse formats.

* Create a standard XML format for several record sets in diverse formats.
* Project team members from the courts expect that the current XML standard will not be able to preserve the manner in which content is presented. For this reason, the project will explore the use of Adobe PDF in conjunction with XML as an interim standard to preserve presentation formatting and as a single format from which to migrate to future formats.
* Generate within or migrate those record sets to the XML/PDF standard.
* Evaluate whether the information or its presentation has changed in any manner that would affect those records' reliability, trustworthiness, or acceptability as legal evidence of those records.
* Revise the standard XML format until content and presentation of the data in the diverse test sets are assessed as reliable and trustworthy.

2. Establish policies and procedures to ensure the authenticity of the electronic records through their life cycle.
   * Within the office of origin during creation and use.
   * Within the office of origin during inactive use.
   * During and after scheduled transfer of the records from the office of origin to a central electronic records storage facility.
   * Within a central electronic records management facility responsible for inactive storage (including providing access to and migrating the records, when necessary, to counter technology obsolescence).
   * For transfer of the records from the central electronic records management center or archive back to the office of origin (or other user).
   * Provide standardized validation process for conversion of records to format for electronic storage and use to insure legal acceptability of the electronic copy of the record.

3. Establish techniques to ensure the authenticity of the electronic records.
   * Investigate the use of electronic signatures, including procedures to preserve access to signature keys during the life cycle of the record.
   * Identify when a document was signed, both when it was created or when it was reviewed along the path to becoming a record.

4. Identify the necessary resources agencies need to implement these policies and procedures.
   * Determine the minimum staff resources and skills necessary to manage electronic records.
   * Determine life cycle costs for managing electronic records to improve planning and budgeting at agency and state records repository.
   * Determine what training the ASLAPR will need to provide records officers.

SIGNIFICANCE AND RELATIONSHIP TO NHPRC GOALS

The NHPRC has identified ten research issues in electronic records. This project addresses several of those issues, including

* The technological, conceptual, and economic implications of capturing and retaining data, descriptive information, and contextual information in electronic form from a variety of applications (Question 2)
* Techniques to retain software-dependent data objects for future use (Question 3)
* The policies that address archival concerns for the identification, retention, preservation, and research use of electronic records (Question 6)

In order to ensure that citizens, state agencies, and others can continue to access electronic records, it is essential that Arizona have effective policies and procedures for creating and managing those records throughout their life cycle. The project seeks to address a number of issues.

Preservation of the Raw Data / Media

Records must be protected from physical damage, whether from deterioration of the media, pestilence, theft, or disaster. Because electronic records managers have developed procedures to overcome problems related to the preservation of the raw electronic data, this grant will not investigate those problems beyond incorporating generally accepted practices into the training materials developed.

Preservation of Meaning

It is essential that individuals who consult records after they are created understand the context and content of the records. Documents and law from the nation's founding fathers can be easily read and understood, although they were written more than 200 years ago. However, the techniques used to record electronic records requires specific hardware, and those records are usually encoded using a manufacturer's proprietary format. Without the appropriate hardware and software, the record may be unreadable. As software applications evolve or become obsolete, it may be impossible to retrieve the content of records. For example, records written in WordStar in 1985 may not be readable by Microsoft Word, and records created in 1995 using MS Word for DOS may not be readable by MS Word 97.

In addition to preserving access to the content (textual information) of the record, the manner in which the content is presented may provide meaningful context for the record. For example, if a record refers to a specific line or page, changes in font, font size, margins, or pagination may change the meaning of the record. Microfilming is an accepted surrogate for paper records, even though the migration to microfilm transforms the presentation of the information; loss of color and contrast may make a document hard to read or make references to color meaningless. Preservation efforts must preserve and retain the contextual information needed to convey the meaning of the information conveyed by the electronic record.

Preservation of Trustworthiness and Reliability

Records need to be accepted as authentic and unaltered documents. Records managers and archivists implement procedures to protect paper records from theft or mutilation, and forgeries and alterations can generally be detected. Electronic records are at greater risk of alteration; in many cases, records custodians may have rights to update and modify original records, and hackers can potentially gain access to servers and delete or alter data. It is extremely difficult to determine if an electronic record has been changed, altered, or forged.

While techniques have been developed to sign or encrypt electronic documents in order to ensure authenticity, they are not widely used by records creators. Further, keys to access signed or encrypted documents must also be protected.

In addition to these general problems of electronic records, four other factors make the timing of this research project critical.

1. Many of the records are stored in proprietary formats and require specific application software to read them. If the application needed to view the record is no longer available, the record becomes unavailable. Few agencies retain copies of old software and hardware, or migrate the electronic records in their care to insure access to the records over time.

2. Arizona, like many other states, is on the verge of accepting electronically signed documents as public records; state law enables state agencies to use electronic records (A.R.S. §44-7041) and recognizes electronic signature use by state agencies (A.R.S. §44-132). Arizona records custodians must find a way to ensure that electronic signatures are reliable, secure, accessible and are as acceptable to records users as are traditional paper documents, especially as legal evidence. At the same time, it is important to ensure that signature keys are preserved, otherwise signatures may themselves become a barrier to access.

3. Many state agencies that create, handle, and keep electronic records lack the resources or the expertise to effectively manage them. Arizona agencies need to know what personnel, hardware, software and operational resources and training are necessary to ensure that records are not lost or compromised through mishandling.

4. The State Library and Archives needs to assess the scope and content of electronic records generated by state government, and to plan for the implications of long-term storage of records received in electronic form.

Records creators, as well as records managers and archivists, are continually confronted with new types of record formats and media. Failure to find solutions to the problems of managing and preserving the data and context, and

successfully migrating to new media, hardware and software formats may mean that many important records will be lost.

This project proposes to investigate the combined use of XML and PDF to provide a standard data format to store, manage, and retrieve records created in diverse applications. Because all records will be contained in a standard format, it will not be necessary to build tools to migrate records from many application-specific formats as each application becomes obsolete. In addition XML schema and metadata tags will capture and organize information used for managing and maintaining records throughout their life cycle.

XML was created and developed by the W3C XML Working Group, which includes key industry players such as Adobe, ArborText, Hewlett-Packard, Isogen, Microsoft, NCSA, Netscape, SoftQuad, and Sun Microsystems (http://www.w3.org/Press/1998/XML10-REC). Many software developers are adopting XML, including Microsoft and Netscape. An Internet Engineering Task Force (IETF) working group is developing standards to electronically sign XML documents. LegalXML, a non-profit organization, is using XML to develop technical standards for legal documents (http://www.legalXML.org/).

XML is a widely used, open standard; it is free from the limitations of proprietary formats and eliminates dependence on a single developer. It is text-based and well-documented; any developer can create applications using XML. Because many developers are using XML, a wide range of tools to help create those applications are available.

PDF is a proprietary format developed by Adobe. While PDF is not an open standard, it has broad support from private, public, and government sectors. The Adobe Acrobat viewer is distributed at no cost and is easily incorporated into both the Microsoft Internet Explorer and Netscape Navigator browsers. The project team views PDF as an interim format to preserve the presentation of information. Software to convert PDF to other formats is available from many vendors and should remain available far beyond the time needed to migrate records to other viable standards in the future.

Records in diverse formats will be migrated to PDF to preserve content and presentation. Records will also be created in XML. Each PDF object and XML object will be embedded in an XML wrapper, which will contain information necessary to locate and manage the record, including agency of origin, series, date, and disposition dates.

Finally, the records will be electronically signed to ensure authenticity. While techniques have been developed to sign electronic documents, those techniques are not widely used. Further, keys to access signed or encrypted documents must also be protected and refreshed.

The project seeks to add to the growing body of research into the problems of electronic records. Similar research includes:

San Diego Supercomputer Center MIX Project. Explored the use of XML views of data exported with a semantic description of the content and interfaces (XML queries) used for viewing the data. (http://www.db.ucsd.edu/projects/MIX). This project looks specifically at e-mail, especially in large quantities. The Arizona project looks at other record types.

State Archives Department, Minnesota Historical Society, Trustworthy Information Systems. Explored tools necessary for the preservation of authenticity and reliability of electronic information systems. (Trustworthy Information Systems Handbook, March 2000). This project looked at preserving confidence in the authenticity of records. The Arizona project will address many related issues of policy and procedure, but it will emphasize the use of an open-standard, platform-neutral data format and the use of electronic signatures to preserve access.

Arizona State University, Electronic Policy Manuals. Explored issues of legal admissibility of the university's policies and procedures manuals, which are now distributed solely as web documents. The university had to

develop policies and procedures to record changes in order to be able to identify and present an authoritative copy of the official policy in effect at any given time.

Michigan Department of Management and Budget, Use of a DoD 5015.2-Certified Records Management Application. This and future Initiative projects will be working with records from several different agencies, many of which are too small to justify a full-fledged records management application.

Rhode Island Office of the Secretary of State, Electronic Records Program Model and Manual. The Arizona project will benefit significantly from the theoretical overview and manual being developed in Rhode Island. It will also produce a manual that will serve as a point of comparison to the Rhode Island manual.

This project will no doubt identify many additional areas for future research. Preliminary work has already identified issues that Arizona needs to address in future projects.

Role of Records Management Centers. The traditional role of storing voluminous, inactive records in less expensive, off-site record centers may become irrelevant, as electronic records take relatively little physical space. However, future centers may find their role is to disseminate standards, coordinate training and planning related to records management for other agencies, provide access to users and originating agencies, and to refresh and migrate data over time.

Control and Custody of Records. Most record centers hold inactive records for other agencies until they are disposed of or transferred to an archive. Records centers may find that electronic records will be managed differently and will likely involve storage of records at remote locations on equipment managed by the creating department as well as in the traditional records center. Similarly, questions to explore include determining if archives will take both title and physical custody of records, or if they will maintain permanent records housed elsewhere.

Storage During Life Cycle. Records are typically stored in the office of the creator during active use, transferred to a records center during inactive use, and then disposed of or transferred to an archive for permanent retention. In the emerging electronic environment, it will be possible for record centers or archives to store a copy of all records from the moment of their creation through replication across servers.

WORK PLAN (see Timeline)

This project will test these policies, procedures, and techniques on three diverse sets of records originally created in electronic format. The project will develop tools to create or migrate the records to a single standard in order to minimize the number of data formats and the need for proprietary software applications to read the data. The records will be transferred to inactive storage at the Records Management Division; access to needed records from the set will be provided to authorized users and to the office of origin. Throughout the process, the records will be evaluated to ensure that they remain accurate, reliable, and trustworthy. Finally, all procedures and policies associated with this process will be documented in a handbook.
This project is expected to take two years, commencing in February 2001 and ending in January 2003.

1. Project Startup and Oversight

The project team will begin by familiarizing all members with all aspects of the problem. Archivists and records managers will gain expertise in information technology, and information technologists will become familiar with archival principles. The team will accomplish this three ways:

* An intensive review of existing literature.
* Attending a workshop on electronic records taught by Thomas Ruller, an expert in electronic archives.
* Send a team member to training programs to ensure that the project has up-to-date knowledge of XML directions and methodologies.
* Send a team member to consult onsite with research projects at San Diego, Ohio, and Minnesota.

* Bring in an advisory board of experts in electronic records to ensure that this project is well-informed about other research initiatives to avoid duplication of effort and to ensure that this project cooperates with other electronic-records research projects. The advisory board will help the project team refine the research agenda and methodology.

Throughout the process, the project management team will meet monthly to ensure that the work remains on track. The project management team will publish quarterly progress reports on the Web to ensure that other researchers are aware of its work and can benefit for what the team has learned.

2. Initial Development of Standard File Format

This project will use a relatively small set of data from the participating agencies to permit an iterative process of development, evaluation and revision. The goal is to lay the groundwork for Arizona to establish non-proprietary methods to standardize electronic records in various forms into common groups that maintain the record's integrity while reducing the variations in software format. This project will work with several common word processing formats (e.g. Microsoft Word, WordPerfect) that can be converted to a common PDF format. The PDF file can then be wrapped within an XML form. The "final" electronic record can then be archived as either an XML document or a PDF document with an XML wrapper. The proposed process will allow the project team to develop practical procedures to convert the sample electronic records into common groups and index them within the common XML envelope to improve long-term management.

The project will also address original electronic documents created and signed within XML. A key question is what different needs are there with these documents which may be organized for on-line viewing of the final document rather than for printing it.

Ensuring the validity of electronic signatures is critically important as documents are migrated through their life cycle. Similarly, it is critical to ensure the veracity of the signature and accompanying document for legal use. The second phase of the project will build and test tools for electronically notarizing signed electronic documents and investigate methods to determine where and how to "refresh" the signature as the technology ages.

After the development of the master XML schema and wrapper, separate DTDs will be developed for three diverse record sets. The process will be iterative, refining earlier DTDs with knowledge gained in later development. The record sets include:

Administrative Office of the Courts

Supreme Court Administrative Orders have been entirely paper-based, even though they have been prepared in electronic word processing formats for the last several years. Currently Administrative Orders are archived on paper by the Clerk of the Supreme Court. A project within the Courts attempts to make these available electronically and to maintain an archival electronic copy that includes digital signature verification.

Administrative Orders are internally prepared directives to courts regarding administrative policies, procedures, committee appointments and other requirements. Order may be superseded by new orders or may expire. Administrative Orders are signed by the Chief Justice and become permanent records of the administrative policies, rules, requirements and procedures in effect at a given time.

Clerk of the Arizona Superior Court in Maricopa County

This data set is similar to the Administrative Office of the Courts. These documents will be electronically filed with the court using PDF with an XML wrapper, but the records will not be electronically signed.

Secretary of State

Recent legislative changes allow the Secretary of State to accept electronic forms of lobbyist's reports. In addition, the Secretary of State will initiate a project this year to build a means for accepting lobbyists' reports, which have been electronically filed and signed. This process will require electronic signatures on XML documents.

Team members from the Courts will work with developers to ensure that procedures and DTDs being developed are practical for use in the agency of origin. Further, they will work with the developers to ensure that the records remain acceptable as legal evidence. A consultant with technical expertise in XML will be brought in to review the XML schema, wrapper, and first DTD before beginning work on the DTD for the second record set.

3. Mid-Project Evaluation

At this point, the advisory board and project team will meet to review work done to date and to refine the final year of development.

4. Test of a Standard File Format

The initial trials will have been done on a Windows-based server at the Secretary of State. Once the initial trials are complete, a second Linux-based XML server will be installed in the ASLAPR Records Management Division. Record sets will be transmitted from the Secretary of State's server to the ASLAPR server to test movement from an active data management system to an off-site electronic records management center. This phase of the project will further test the XML schema, DTDs, and the benefits of open-standards by simulating transfer of records from a server in the agency of origin to another server running a different operating system and XML server software.

The ASLAPR Records Management Division will then simulate long-term storage of records in an off-site electronic records management center. The ability to locate, retrieve, and return specific records with accuracy and reliability will complete the proof of concept.

5. Evaluation

A consultant with technical expertise in XML will review the XML wrapper, schema, and DTDs, and will help develop evaluation criteria to test the policies, procedures, and methodologies developed during the project. State librarians from Alaska, Connecticut, Florida, Kentucky, Rhode Island, Texas, and Virginia have agreed to help evaluate the tools to ensure that they will be useful outside Arizona.

6. Documentation of Policies and Procedures

Throughout the development process, ASLAPR staff will ensure that the integrity of the records is not compromised. They will advise the developers as to when certifications are appropriate and watch for procedures that may place the records at risk. ASLAPR staff will document procedures as they are developed.

The project staff will work with a professional technical writer to incorporate the policies and procedures into a manual. That manual will also include instructions for use of the software tools. ASLAPR will be responsible for producing and printing the manual and for production of an online version.

PRODUCTS AND PUBLICATIONS

The project will establish and test policies and procedures and an open-standards based format for storing and managing disparate electronic records. Creating and testing this mechanism will lead to defining methods and common formats to standardize various forms of electronic records The goal is to determine an effective means for migrating and archiving these electronic records to insure current and future access and usability.

The published materials that result from this project will include

* Quarterly progress reports on the web
* A final report on the web
* A policy and procedures manual

Products produced will include:

* Tools and practices to convert word processing documents to PDF format
* Tools and practices to create electronic signatures on XML documents

* Tools and practices to create XML schema, containers, and documents
* Conversion tools and practices to “refresh” signed electronic documents

KEY PERSONNEL

Project Manager:

Russ Savage, Electronic Signatures Liaison, Secretary of State. Coordinates all activities, chair of the Project Management Team. Advises on electronic signatures.

Project Management Team

Michael Totherow, Chief Information Officer, Secretary of State. Supervises staff and contract programmers.
GladysAnn Wells, State Librarian, ASLAPR. Advises on state information policy issues.
David Hoober, Director, ASLAPR Archives and History Division. Advises on archival practice.
Martin Richelsoph, Director, ASLAPR Records Management Division. Advises on records management practice, writes manual.
Richard Pearce-Moses, Coordinator, ASLAPR Cultural Inventory Project. Advises on archival practice.
Maureen Haggerty, Manager, Information Technology Division, Arizona Administrative Office of the Courts. Coordinates and provides technical advice for court records.
Ted Wilson. Courts Services Specialist VII, Arizona Administrative Office of the Courts. Provides advice for court records.
Jeremy Rowe, Head, ASU Media Development Information Technology. Advises on XML, PDF.

Project Support Staff

Jason Wharton, Systems Analyst/Programmer III, Secretary of State. Helps develop conversion tools.
Michael Fasciano, Systems Analyst/Programmer III, Secretary of State. Helps develop conversion tools.
Peggy Zeller, Graphic designer, ASLAPR. Markup and production of manuals.
Carol Westwood, CFO, ASLAPR. Financial management.

Advisory Board

Anne Gilliland-Swetland, Assistant Professor, Dept. of Information Studies, University of California, Los Angeles
Eliot Christian, United States Geological Survey
Phil Coombs, State Archivist, Washington State
Shawn Rounds, Government Records Specialist, Minnesota State Archives, Minnesota Historical Society
Judy Walker, Electronic Records Archivist, Ohio Historical Society

Contract Labor

Thomas Ruller. Teaches workshop on electronic records.
XML consultant. To be named. Provides expertise and XML and digital signatures
Contract programmers. To be named. Develops migration tools.
Technical writer. To be named. Edits policies/procedures manual.

State Library Evaluators

Alaska State Library Archives and Museums
Connecticut State Library
Florida Division of Library and Information Services
Kentucky Department for Libraries and Archives
Rhode Island Office of Library and Information Services
Texas State Library and Archives Commission
Library of Virginia